MARYVILLE UNIVERSITY

Introduction of

Week of 9/11/17

Part C

MARYVILLE UNIVERSITY

R Introduction

R 101

Input Data

Build a model

Test/Validation/Prediction

Output/Graphics

R Introduction

MARYVILLE UNIVERSITY

1. Open source software application for statistical computing & graphics http://www.r-project.org
2. A brief history
   - S programming language, developed at Bell Lab in 70s
   - By ***R***oss Ihaka & **R**obert Gentleman at U. of Auckland, NZ
   - S-PLUS is commercial version
3. Advantage

   Comprehensive: 7,000+ add-on packages & 120K+ functions(08/15/16); Modern statistical language, close to C/C++; Advanced graphic capability; Large online community; Cross-plateform applications; Integrated with many tools; **Free**, no license restriction
4. Disadvantage

   Steep learning curve; Memory demand; Single thread; Command line driven (limited GUI); No technical support; No liability; (**Free**)

R 101

MARYVILLE UNIVERSITY

- Run & close R
- Assign value to a variable (= or <-)
- Display value (variable or (…))
- Run command from script editor
- Retrieve historical command (up/down arrow)
- Comment line (start from #)
- Workspace; objects (list, remove)
- Built-in help system; online community
- Use R as number calculator
- All about vector in R; vectorization
- Sequence and replicate

Data Input & Manipulation

MARYVILLE UNIVERSITY

R can import data from different sources in different format;

- File, Microsoft Excel/Access/SQL, Oracle, MySQL, HTTP, etc.
- CSV, Excel, table, SAS, Minitab, SPSS, XML, etc.

Most common: CSV or text file (read.csv, read.table, etc.; scan)
Data.frame structure
Access to data frame & manipulation

Use R existing data

MARYVILLE UNIVERSITY

Build a model

Linear model – lm; use online help system to find details and examples

Formula & its options

Other options – data, weights

Look at modeling results – summary

MARYVILLE UNIVERSITY

Model test/validation/prediction

Summary of a model

Test on additional variables (adj. R^2, F-test, AIC)

Validation – covered in future class

Prediction – in-sample & out-sample

MARYVILLE UNIVERSITY

Output Results

Make plots out of model

Output plots to file & clipboard

Write model results to files

Graphic examples

PCH option